digest of papers

THE AMPERIF CACHE DISK SYSTEM

STEPHEN FULD

AMPERIF CORPORATION

The Amperif Cache Disk System provides high performance, large capacity, very flexible, low cost of ownership storage for users of a variety of mainframe CPUs. High speed logic combined with full read and write, track oriented cache, multiple channel ports and data paths as well as disk array design give a high performance level. The modular architecture of the system provides flexibility for a variety of host interfaces, storage configurations and operating modes. The system uses a fraction of the floor space, power and air conditioning of comparable products. Of course, due care was taken to assure data integrity throughout the system.

The Amperif Cache Disk System is a large capacity, high performance, caching disk controller for selected mainframe environemnts. It has been adapted for a variety of interfaces and a variety of configurations. This paper gives an overview of the system and explains how the design elements contribute to high performance, flexability, low cost of ownership and data integrity.

High Performance

Several design features combine to give a high performance system. The system caches full disk tracks. This allows the caching operation to begin anywhere on the track and complete in one disk revolution, effectively eliminating the rotational latency and possible RPS misses of sector oriented caches. The track organization makes the number of cache entries smaller than a sector cache for any given cache size. This allows the implementation of high speed cache search hardware. The hardware performs searches and keeps the pointer stack in LRU order at the rate of 1.25 million entries per second. Thus the typical search time is 20 to 30 microseconds. Future versions will increase the rate to 5 million entries per second.

The tri-bus structure of the controller also contributes to high performance. One bus is dedicated to data on a non-contention, burst basis. This bus can handle a data rate of 20 megabytes per second. Two additional busses are used to send instructions from the main microprocessor to each of the other boards in the system, and to send control parameters and receive status from the other boards. The main microprocessor uses bit slice technology executing one instruction every 200 nanoseconds. Even this isn't fast enough to handle data so all the data paths are under direct hardware control. The basic protocol for a data transfer is for the main microprocessor to tell the sending board to put data on the bus, the receiving board to take it; then initiate both boards and wait for a completion status.

Another factor that contributes to high performance is the use of dual data control paths and up to eight channel interfaces. A request can be sent to any one of the eight and be processed by whichever of the two data control paths isn't busy. This reduces control unit busy time.

Flexibility

Because of the building block structuring of the Amperif Cache Disk System, components can be changed and intermixed in a variety of ways. For example, the system supports up to 32 logical units, each of which can be disk, cache disk, or solid state storage. The disks available span a wide capacity range allowing up to 32 gigabytes of storage on a string. Optionally, two disks can be treated as one by the microcode to allow larger capacities on systems where the number of units is a constraint.

For each cache disk unit, there are three operating modes. Bypass mode allows a cache disk unit to function as a non cache unit; all operations bypass the cache and go directly to/from the disk. Write thru allows caching with utmost assurance of data integrity. Read

operations are handled from the cache and write operations are written to both the cache and the disk. The completion status is delayed until successful completion of the disk write. Full or post store cache is the same as write thru except the completion status for writes is given after the data is in cache. This provides the maximum performance mode of the system. Each unit's operating mode can be changed, concurrently with full operations, either by the operator or through system commands.

The system supports a variety of host computer interfaces. The ones presently existing or in development include Unisys (Sperry) word channel, IBM block multiplexor (FBA mode), CDC Cyber channel, Unisys (Burroughs) DLP, and IPI-3. Thus a user can protect his investment by buying a system that operates with his current and possible future computers. In some cases, multiple interfaces can be mixed on the same controller for flexibility in multi-vendor shops.

Low Cost of Ownership

By using a disk array design, the system reduces cost of ownership. For one thing, it uses very little floor space. A system with two controllers, all associated interface hardware, 16 gigabytes of disk, 32 megabytes of cache, and 45 megabytes of solid state storage fits in one 12 square foot cabinet. An additional 16 gigabytes of disk takes only 6 square feet. Another benefit of the disk array concept is the reduction in power and air conditioning required. Typically, the Amperif system uses about one half the power of competing systems.

Data Integrity

The high performance, flexibility, and low cost are worthless if the disk doesn't do the one thing that you expect it to do; store and transfer data without losing or gaining even a single bit. Features are designed into the Amperif Cache Disk System to assure the data integrity that mainframe users require. For cache and solid state systems, there is a UPS that prevents data loss in a power outage. Additionally, to handle the cases where the 15 minutes provided by the UPS isn't sufficient, there is a data save disk. If the power is out for more than 1 minute, the contents of the cache and Solid State Disk are copied to the data save disk. Then, no matter how long the outage is, when the power comes on again, the data can be copied back to the cache and Solid State Disk from the data save disk. Thus no data will be lost.

The controllers are designed to high standards of hardware integrity. There is parity on all internal busses. Data is through checked to prevent a part failure from causing undetected data corruption. Both the disk and the memory use modern ECC algorithms. If the controller sees correctable ECC errors on the memory, it corrects that data, then dynamically downs that block of memory in order to prevent any future uncorrectable errors. The critical block number to cylinder, head and record number division is checked by multiplying the numbers back and testing for the original number. By these means, the rate of undetected data corruption is reduced to one in 50 years.

By combining a modular design with intelligent caching strategies and using disk arrays, the Amperif Cache Disk System provides an alternative to the storage typically provided by mainframe vendors.